

Information geometric complexity of a trivariate Gaussian statistical model

Domenico Felice

School of Science and Technology

University of Camerino, I-62032 Camerino, Italy

INFN-Sezione di Perugia, Via A. Pascoli, I-06123 Perugia, Italy

Carlo Cafaro

Department of Mathematics

Clarkson University, Potsdam, 13699 NY, USA

Stefano Mancini

School of Science and Technology

University of Camerino, I-62032 Camerino, Italy

INFN-Sezione di Perugia, Via A. Pascoli, I-06123 Perugia, Italy

April 1, 2014

Abstract

We evaluate the information geometric complexity of entropic motion on low-dimensional Gaussian statistical manifolds in order to quantify how difficult is making macroscopic predictions about a systems in the presence of limited information. Specifically, we observe that the complexity of such entropic inferences not only depends on the amount of available pieces of information but also on the manner in which such pieces are correlated. Finally, we uncover that for certain correlational structures, the impossibility of reaching the most favorable configuration from an entropic inference viewpoint, seems to lead to an information geometric analog of the well-known frustration effect that occurs in statistical physics.

Keywords: Probability theory; Riemannian geometry; Complexity

1 Introduction

One of the main efforts in physics is modelling and predicting natural phenomena using relevant information about the system under consideration. Theoretical physics has had a general measure of the uncertainty associated with the behaviour of a probabilistic process for more than 100 years: the Shannon entropy [1]. The Shannon information theory was applied to dynamical systems and became successful in describing their unpredictability [2].

Along a similar avenue we may set Entropic Dynamics [3] which makes use of inductive inference (Maximum Entropy Methods [4]) and Information Geometry

[5]. This is clearly remarkable when microscopic dynamics can be too far removed from the phenomena of interest, such as in complex biological or ecological systems. Extension of ED to temporally-complex dynamical systems on curved statistical manifolds brought to the finding of relevant measures of chaoticity [6]. In particular, an Information Geometrodynamical Approach to Chaos (IGAC) has been pursued studying chaos in informational geodesic flows describing physical, biological or chemical systems. It is the information geometric analogue of conventional geometrodynamical approaches [7] where the classical configuration space is being replaced by a statistical manifold with the additional possibility of considering chaotic dynamics arising from non conformally flat metrics. Within this framework, it looks like natural to consider as complexity measure the (time average) statistical volume explored by geodesic flows, namely an Information Geometry Complexity (IGC).

This quantity might help in uncovering connections between microscopic dynamics and experimentally observable macroscopic dynamics which is a fundamental issue in physics [8]. An interesting manifestation of such a relationship appears in the study of the effects of microscopic external noise (noise imposed on the microscopic variables of the system) on the observed collective motion (macroscopic variables) of a globally coupled map [9]. These effects are quantified in terms of the complexity of the collective motion. Furthermore, it turns out that noise at a microscopic level reduces the complexity of the macroscopic motion, which in turn is characterized by the number of effective degrees of freedom of the system.

The investigation of the macroscopic behaviour of complex systems in terms of the underlying statistical structure of its microscopic degrees of freedom also reveals effects due to the presence of microcorrelations [10]. In this article we first show which macro-states should be considered in a Gaussian statistical model in order to have a reduction in time of the Information Geometry Complexity. Then, dealing with correlated bivariate and trivariate Gaussian statistical model, the ratio between the IGC in the presence and in the absence of microcorrelations is explicitly computed, finding an intriguing, even though non yet deep understood, connection with the phenomenon of geometric frustration [11].

The layout of the article is as follows. In Section 2 we introduce a general statistical model discussing its geometry and describing both its dynamics and information geometry complexity. In Section 3, Gaussian statistical models (up to a trivariate model) are considered. There, we compute the asymptotic temporal behaviours of their IGCs. Finally, in Section 4 we draw our conclusions by outlining our findings and proposing possible further investigations.

2 Statistical models and information geometry complexity

Given n real-valued random variables X_1, \dots, X_n defined on the sample space Ω with joint probability density $p : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfying the conditions

$$p(x) \geq 0 \ (\forall x \in \mathbb{R}^n) \quad \text{and} \quad \int_{\mathbb{R}^n} dx \ p(x) = 1,$$

let us consider a family \mathcal{P} of such distributions and suppose that they can be parametrized using m real-valued variables $(\theta^1, \dots, \theta^m)$ so that

$$\mathcal{P} = \{p_\theta = p(x|\theta) | \theta = (\theta^1, \dots, \theta^m) \in \Theta\},$$

where $\Theta \subseteq \mathbb{R}^m$ is the parameter space and the mapping $\theta \rightarrow p_\theta$ is injective. In such a way, \mathcal{P} is an m -dimensional statistical model on \mathbb{R}^n .

The mapping $\varphi : \mathcal{P} \rightarrow \mathbb{R}^m$ defined by $\varphi(p_\theta) = \theta$ allows us to consider $\varphi = [\theta^i]$ as a coordinate system for \mathcal{P} . Assuming parametrizations which are C^∞ , we can turn \mathcal{P} into a C^∞ differentiable manifold (thus, \mathcal{P} is called statistical manifold) [5].

The values x_1, \dots, x_n taken by the random variables define the *micro-state* of the system, while the values $\theta^1, \dots, \theta^m$ taken by parameters define the *macro-state* of the system.

Let $\mathcal{P} = \{p_\theta | \theta \in \Theta\}$ be an m -dimensional statistical model. Given a point θ , the Fisher information matrix of \mathcal{P} in θ is the $m \times m$ matrix $G(\theta) = [g_{ij}]$, where the (i, j) entry is defined by

$$g_{ij}(\theta) := \int_{\mathbb{R}^n} dx p(x|\theta) \partial_i \log p(x|\theta) \partial_j \log p(x|\theta), \quad (1)$$

with ∂_i standing for $\frac{\partial}{\partial \theta^i}$. The matrix $G(\theta)$ is symmetric, positive semidefinite and determines a Riemannian metric on the parameter space Θ [5]. Hence, it is possible to define a Riemannian statistical manifold $\mathcal{M} := (\Theta, g)$, where $g = g_{ij} d\theta^i \otimes d\theta^j$ ($i, j = 1, \dots, m$) is the metric whose components g_{ij} are given by Eq. (1) (throughout the paper we use the Einstein sum convention).

Given the Riemannian manifold $\mathcal{M} = (\Theta, g)$, it is well known that there exists only one linear connection ∇ (the Levi-Civita connection) on \mathcal{M} that is compatible with the metric g and symmetric [12]. We remark that the manifold \mathcal{M} has one chart, being Θ an open set of \mathbb{R}^m , and the Levi-Civita connection is uniquely defined by means of the Christoffel coefficients

$$\Gamma_{ij}^k = \frac{1}{2} g^{kl} \left(\frac{\partial g_{lj}}{\partial \theta^i} + \frac{\partial g_{il}}{\partial \theta^j} - \frac{\partial g_{ij}}{\partial \theta^l} \right), \quad (i, j, k = 1, \dots, m) \quad (2)$$

where g^{kl} is the (k, l) entry of the inverse of the Fisher matrix $G(\theta)$.

The idea of curvature is the fundamental tool to understand the geometry of the manifold $\mathcal{M} = (\Theta, g)$. Actually, it is the basic geometric invariant and the intrinsic way to obtain it is by means of geodesics. It is well-known, that given any point $\theta \in \mathcal{M}$ and any vector v tangent to \mathcal{M} at θ , there is a unique geodesic starting at θ with initial tangent vector v . Indeed, within the considered coordinate system, the geodesics are solutions of the following nonlinear second order coupled ordinary differential equations [12]

$$\frac{d^2 \theta^k}{d\tau^2} + \Gamma_{ij}^k \frac{d\theta^i}{d\tau} \frac{d\theta^j}{d\tau} = 0, \quad (3)$$

with τ denoting the time.

The recipe to compute some curvatures at a point $\theta \in \mathcal{M}$ is the following: first, select a 2-dimensional subspace Π of the tangent space to \mathcal{M} at θ ; second, follow the geodesics through θ whose initial tangent vectors lie in Π and consider the 2-dimensional submanifolds S_Π swiped out by them inheriting a Riemannian metric

from \mathcal{M} ; finally, compute the Gaussian curvature of S_Π at θ , which can be obtained from its Riemannian metric as stated in the *Theorema Egregium* [13]. The number $K(\Pi)$ found in such manner is called the *sectional curvature* of \mathcal{M} at θ associated with the plane Π . In terms of local coordinates, to compute the sectional curvature we need the curvature tensor,

$$R_{ijk}^h = \frac{\partial \Gamma_{jk}^h}{\partial \theta^i} - \frac{\partial \Gamma_{ik}^h}{\partial \theta^j} + \Gamma_{jk}^l \Gamma_{il}^h - \Gamma_{ik}^l \Gamma_{jl}^h. \quad (4)$$

For any basis (ξ, η) for a 2-plane $\Pi \subset T_\theta \mathcal{M}$, the sectional curvature at $\theta \in \mathcal{M}$ is given by [12]

$$K(\xi, \eta) = \frac{R(\xi, \eta, \eta, \xi)}{|\xi|^2 |\eta|^2 - \langle \xi, \eta \rangle^2}, \quad (5)$$

where R is the Riemann curvature tensor which is written in coordinates as $R = R_{ijkl} d\theta^i \otimes d\theta^j \otimes d\theta^k \otimes d\theta^l$ with $R_{ijkl} = g_{lh} R_{ijk}^h$ and $\langle \cdot, \cdot \rangle$ is the inner product defined by the metric g .

The sectional curvature is directly related to the topology of the manifold; along this direction the *Cartan-Hadamard* Theorem [13] is enlightening by stating that any complete, simply connected n -dimensional manifold with non positive sectional curvature is diffeomorphic to \mathbb{R}^n .

We can consider upon the statistical manifold $\mathcal{M} = (\Theta, g)$ the macro-variables θ as accessible information and then derive the information dynamical equation (3) from a standard principle of least action of Jacobi type [3]. The geodesic equations (3) describe a reversible dynamics whose solution is the trajectory between an initial and a final macrostate θ^{initial} and θ^{final} , respectively. The trajectory can be equally traversed in both directions [10]. Actually, an equation relating instability with geometry exists and it makes hope that some global information about the average degree of instability (chaos) of the dynamics is encoded in global properties of the statistical manifolds [7]. The fact that this might happen is proved by the special case of constant-curvature manifolds, for which the Jacobi-Levi-Civita equation simplifies to [7]

$$\frac{d^2 J^i}{d\tau^2} + K J^i = 0, \quad (6)$$

where K is the constant sectional curvature of the manifold (see (5)) and J is the geodesic deviation vector field. On a positively curved manifold, the norm of the separating vector J does not grow, whereas on a negatively curved manifold, the norm of J grows exponentially in time, and if the manifold is compact, so that its geodesic are sooner or later obliged to fold, this provide an example of chaotic geodesic motion [14].

Taking into consideration these facts, we single out as suitable indicator of dynamical (temporal) complexity, the information geometric complexity defined as the average dynamical statistical volume [15]

$$\widetilde{\text{vol}}[\mathcal{D}_\Theta^{(\text{geodesic})}(\tau)] := \lim_{\tau \rightarrow \infty} \left(\frac{1}{\tau} \int_0^\tau d\tau' \text{vol}[\mathcal{D}_\Theta^{(\text{geodesic})}(\tau')] \right), \quad (7)$$

where

$$\text{vol}[\mathcal{D}_\Theta^{(\text{geodesic})}(\tau')] := \int_{\mathcal{D}_\Theta^{(\text{geodesic})}(\tau')} \sqrt{\det(G(\theta))} d\theta, \quad (8)$$

with $G(\theta)$ the information matrix whose components are given by Eq. (1). The integration space $\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau')$ is defined as follows

$$\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau') := \{\theta = (\theta^1, \dots, \theta^m) : \theta^k(0) \leq \theta^k \leq \theta^k(\tau')\}, \quad (9)$$

where $\theta^k \equiv \theta^k(s)$ with $0 \leq s \leq \tau'$ such that $\theta^k(s)$ satisfies (3). The quantity $\text{vol}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau')]$ is the volume of the effective parameter space explored by the system at time τ' . The temporal average has been introduced in order to average out the possibly very complex fine details of the entropic dynamical description of the system's complexity dynamics.

Relevant properties, concerning complexity of geodesic paths on curved statistical manifolds, of the quantity (8) compared to the Jacobi vector field are discussed in Ref.[16].

3 The Gaussian statistical model

In the following we devote our attention to a Gaussian statistical model \mathcal{P} whose element are multivariate normal joint distributions for n real-valued variables X_1, \dots, X_n given by

$$p(x|\theta) = \frac{1}{\sqrt{(2\pi)^n \det C}} \exp \left[-\frac{1}{2}(x - \mu)^t C^{-1}(x - \mu) \right], \quad (10)$$

where $\mu = (\mathbb{E}(X_1), \dots, \mathbb{E}(X_n))$ is the n -dimensional mean vector and C denotes the $n \times n$ covariance matrix with entries $c_{ij} = \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j)$, $i, j = 1, \dots, n$. Since μ is a n -dimensional real vector and C is a $n \times n$ symmetric matrix, the parameters involved in this model should be $n + \frac{n(n+1)}{2}$. Moreover C is a symmetric, positive definite matrix, hence we have the parameter space given by

$$\Theta := \{(\mu, C) | \mu \in \mathbb{R}^n, C \in \mathbb{R}^{n \times n}, C > 0\}. \quad (11)$$

Hereafter we consider the statistical model given by Eq. (10) when the covariance matrix C has only variances $\sigma_i^2 = \mathbb{E}(X_i^2) - (\mathbb{E}(X_i))^2$ as parameters. In fact we assume that the non diagonal entry (i, j) of the covariance matrix C equals $\rho \sigma_i \sigma_j$ with $\rho \in \mathbb{R}$ quantifying the degree of correlation.

We may further notice that the function $f_{ij}(x) := \partial_i \log p(x|\theta) \partial_j \log p(x|\theta)$, when $p(x|\theta)$ is given by (10), is a polynomial in the variables x_i ($i = 1, \dots, n$) whose degree is not greater than four. Indeed, we have that

$$\partial_i \log p(x|\theta) = \frac{1}{p(x|\theta)} \partial_i p(x|\theta) = \partial_i \frac{1}{\sqrt{(2\pi)^n \det C}} + \partial_i \left[-\frac{1}{2}(x - \mu)^t C^{-1}(x - \mu) \right],$$

and, therefore, the differentiation does not affect variables x_i . With this in mind, in order to compute the integral in (1), we can use the following formula [17]

$$\frac{1}{\sqrt{(2\pi)^n \det C}} \int dx f_{ij}(x) \exp \left[-\frac{1}{2}(x - \mu)^t C^{-1}(x - \mu) \right] = \exp \left[\frac{1}{2} \sum_{h,k=1}^n c_{hk} \frac{\partial}{\partial x_h} \frac{\partial}{\partial x_k} \right] f_{ij}|_{x=\mu}, \quad (12)$$

where the exponential denotes the power series over its argument (the differential operator).

3.1 The monivariate Gaussian statistical model

We now start to apply the concepts of the previous section to a Gaussian statistical model of Eq.(10) for $n = 1$. In this case, the dimension of the statistical Riemannian manifold $\mathcal{M} = (\Theta, g)$ is at most two. Indeed, to describe elements of the statistical model \mathcal{P} given by Eq. (10), we basically need the mean $\mu = \mathbb{E}(X)$ and variance $\sigma^2 = \mathbb{E}(X - \mu)^2$. We deal separately with the cases when the monivariate model has only μ as macro-variable (Case 1), when σ is the unique macro-variable (Case 2), and finally when both μ and σ are macro-variables (Case 3).

3.1.1 Case 1

Consider the monivariate model with only μ as macro-variable by setting $\sigma = 1$. In this case the manifold \mathcal{M} is trivially the real *flat* straight line, since $\mu \in (-\infty, +\infty)$. Indeed, the integral in (1) is equal to 1 when the distribution $p(x|\theta)$ reads as $p(x|\mu) = \frac{\exp\left[-\frac{1}{2}(x-\mu)^2\right]}{\sqrt{2\pi}}$; so the metric is $g = d\mu^2$. Furthermore, from Eqs. (2),(3) the information dynamics is described by the geodesic $\mu(\tau) = A_1\tau + A_2$, where $A_1, A_2 \in \mathbb{R}$. Hence, the volume of Eq.(8) results $\text{vol}\left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau')\right] = \int d\mu = A_1\tau + A_2$; since this quantity must be positive we assume $A_1, A_2 > 0$. Finally, the behaviour of the IGC (7) is

$$\widetilde{\text{vol}}\left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)\right] \approx \left(\frac{A_1}{2}\right)\tau. \quad (13)$$

This shows that the complexity linearly increases in time meaning that acquiring information about μ and updating it, is not enough to increase our knowledge about the micro state of the system.

3.1.2 Case 2

Consider now the monivariate Gaussian statistical model of Eq.(10) when $\mu = \mathbb{E}(X) = 0$ and the macro-variable is only σ . In this case the probability distribution function reads $p(x|\sigma) = \frac{\exp\left[-\frac{x^2}{2\sigma^2}\right]}{\sqrt{2\pi}\sigma}$ while the Fisher-Rao metric becomes $g = \frac{2}{\sigma^2}d\sigma^2$. Emphasizing that also in this case the manifold is flat as well, we derive the information dynamics by means of Eqs.(2) and (3) and we obtain the geodesic $\sigma(\tau) = A_1 \exp[A_2\tau]$. The volume in Eq.(8) then results

$$\text{vol}\left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau')\right] = \int \frac{\sqrt{2}}{\sigma}d\sigma = \sqrt{2}\log[A_1 \exp[A_2\tau]].$$

Again, to have positive volume we have to assume $A_1, A_2 > 0$. Finally, the IGC(7) becomes

$$\widetilde{\text{vol}}\left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)\right] \approx \left(\frac{\sqrt{2}A_2}{2}\right)\tau. \quad (14)$$

This shows that also in this case the complexity linearly increases in time meaning that acquiring information about σ and updating it, is not enough to increase our knowledge about the micro-state of the system.

3.1.3 Case 3

The take home message of the previous cases is that we have to account for both mean μ and variance σ as macro-variables to look for possible non increasing complexity. Hence, consider the probability distribution function is given by,

$$p(x_1, x_2 | \mu, \sigma) = \frac{\exp \left[-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2} \right]}{\sigma \sqrt{2\pi}}. \quad (15)$$

The dimension of the Riemannian manifold $\mathcal{M} = (\Theta, g)$ is two, where the parameter space Θ is given by $\Theta = \{(\mu, \sigma) | \mu \in (-\infty, +\infty), \sigma > 0\}$ and the Fisher-Rao metric reads as $g = \frac{1}{\sigma^2} d\mu^2 + \frac{2}{\sigma^2} d\sigma^2$. Here, the sectional curvature given by Eq.(5) is a negative function and despite the fact that is not constant, we expect a decreasing behaviour in time of the IGC. Thanks to Eq. (2), we find that the only non negative Christoffel coefficients are $\Gamma_{12}^1 = -\frac{1}{\sigma}$, $\Gamma_{11}^2 = \frac{1}{2\sigma}$ and $\Gamma_{22}^2 = -\frac{1}{\sigma}$. Substituting them into Eq.(3) we derive the following geodesic equations

$$\begin{cases} \frac{d^2 \mu(\tau)}{d\tau^2} - \frac{2}{\sigma} \frac{d\sigma}{d\tau} \frac{d\mu}{d\tau} = 0, \\ \frac{d^2 \sigma(\tau)}{d\tau^2} - \frac{1}{\sigma} \left(\frac{d\sigma}{d\tau} \right)^2 + \frac{1}{2\sigma} \left(\frac{d\mu}{d\tau} \right)^2 = 0. \end{cases} \quad (16)$$

The integration of the above coupled differential equations is non-trivial. We follow the method described in [10] and arrive at

$$\sigma(\tau) = \frac{2\sigma_0 \exp \left[\frac{\sigma_0 |A_1|}{\sqrt{2}} \tau \right]}{1 + \exp \left[\frac{2\sigma_0 |A_1|}{\sqrt{2}} \tau \right]}, \quad \mu(\tau) = -\frac{2\sigma_0 \sqrt{2} A_1}{|A_1| \left(1 + \exp \left[\frac{2\sigma_0 |A_1|}{\sqrt{2}} \tau \right] \right)}, \quad (17)$$

where σ_0 and A_1 are real constants. Then, using (17), the volume of Eq.(8) results

$$vol \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau') \right] = \int \frac{\sqrt{2}}{\sigma^2} d\sigma d\mu = \frac{\sqrt{2} A_1}{|A_1|} \exp \left[-\frac{\sigma_0 |A_1|}{\sqrt{2}} \tau \right]. \quad (18)$$

Since the last quantity must be positive, we assume $A_1 > 0$. Finally, employing the above expression into Eq. (7) we arrive at

$$\widetilde{vol} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau) \right] \approx \left(\frac{2}{\sigma_0 A_1} \right) \frac{1}{\tau}. \quad (19)$$

We can now see a reduction in time of the complexity meaning that acquiring information about both μ and σ and updating them allows us to increase our knowledge about the micro state of the system.

Hence, comparing Eqs. (13),(14) and (19) we conclude that the entropic inferences on a Gaussian distributed micro-variable is carried out in a more efficient manner when both its mean and the variance in the form of information constraints are available. Macroscopic predictions when only one of these pieces of information are available are more complex.

3.2 Bivariate Gaussian statistical model

Consider now the Gaussian statistical model \mathcal{P} of the Eq. (10) when $n = 2$. In this case the dimension of the Riemannian manifold $\mathcal{M} = (\Theta, g)$ is at most four. From the analysis of the monivariate Gaussian model in Section 3.1 we have understood that both mean and variance should be considered. Hence the minimal assumption is to consider $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mu$ and $\mathbb{E}(X_1 - \mu)^2 = \mathbb{E}(X_2 - \mu)^2 = \sigma^2$. Furthermore, in this case we have also to take into account the possible presence of (micro) correlations, which appear at the level of macro-states as off-diagonal terms in the covariance matrix. In short, this implies considering the following probability distribution function

$$p(x_1, x_2 | \mu, \sigma) = \frac{\exp \left[-\frac{1}{2\sigma^2(1-\rho^2)} \left((x_1 - \mu)^2 - 2\rho(x_1 - \mu)(x_2 - \mu) + (x_2 - \mu)^2 \right) \right]}{2\pi\sigma^2\sqrt{1-\rho^2}}, \quad (20)$$

where $\rho \in (-1, 1)$.

Thanks to Eq. (12) we compute the Fisher-Information matrix G and find $g = g_{11}d\mu^2 + g_{22}d\sigma^2$ with,

$$g_{11} = \frac{2}{\sigma^2(\rho+1)}; \quad g_{22} = \frac{4}{\sigma^2}. \quad (21)$$

The only non trivial Christoffel coefficients (2) are $\Gamma_{12}^1 = -\frac{1}{\sigma}$, $\Gamma_{11}^2 = \frac{1}{2\sigma(\rho+1)}$ and $\Gamma_{22}^2 = -\frac{1}{\sigma}$. In this case as well, the sectional curvature (Eq.(5)) of the manifold \mathcal{M} is a negative function and so we may expect a decreasing asymptotic behaviour for the IGC. From Eq.(3) it follows that the geodesic equations are,

$$\begin{cases} \frac{d^2\mu(\tau)}{d\tau^2} - \frac{2}{\sigma} \frac{d\sigma}{d\tau} \frac{d\mu}{d\tau} = 0 \\ \frac{d^2\sigma(\tau)}{d\tau^2} - \frac{1}{\sigma} \left(\frac{d\sigma}{d\tau} \right)^2 + \frac{1}{2(1+\rho)\sigma} \left(\frac{d\mu}{d\tau} \right)^2 = 0, \end{cases} \quad (22)$$

whose solutions are,

$$\sigma(\tau) = \frac{2\sigma_0 \exp \left[\frac{\sigma_0|A_1|}{\sqrt{2(1+\rho)}} \tau \right]}{1 + \exp \left[\frac{2\sigma_0|A_1|}{\sqrt{2(1+\rho)}} \tau \right]}, \quad \mu(\tau) = -\frac{2\sigma_0\sqrt{2(1+\rho)}A_1}{|A_1| \left(1 + \exp \left[\frac{2\sigma_0|A_1|}{\sqrt{2(1+\rho)}} \tau \right] \right)}. \quad (23)$$

Using (23) in Eq. (8) gives the volume,

$$\text{vol} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau') \right] = \int \frac{2\sqrt{2}}{\sqrt{1+\rho} \sigma^2} d\sigma d\mu = \frac{4A_1}{|A_1|} \exp \left[-\frac{\sigma_0|A_1|}{\sqrt{2(1+\rho)}} \tau \right]. \quad (24)$$

To have it positive we have to assume $A_1 > 0$. Finally, employing (24) in (7) leads to the IGC,

$$\widetilde{\text{vol}} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau) \right] \approx \left(\frac{4\sqrt{2}}{\sigma_0 A_1} \right) \frac{\sqrt{1+\rho}}{\tau}, \quad (25)$$

with $\rho \in (-1, 1)$. We may compare the asymptotic expression of the ICGs in the presence and in the absence of correlations, obtaining

$$R_{\text{bivariate}}^{\text{strong}}(\rho) := \frac{\widetilde{\text{vol}}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)]}{\widetilde{\text{vol}}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)]_{\rho=0}} = \sqrt{1 + \rho}, \quad (26)$$

where “strong” stands for the fully connected lattice underlying the micro-variables. The ratio $R_{\text{bivariate}}^{\text{strong}}(\rho)$ results a *monotonic* increasing function of ρ .

While the temporal behaviour of the IGC (25) is similar to the IGC in (19), here correlations play a fundamental role. From Eq. (26), we conclude that entropic inferences on two Gaussian distributed micro-variables on a fully connected lattice is carried out in a more efficient manner when the two micro-variables are negatively correlated. Instead, when such micro-variables are positively correlated, macroscopic predictions become more complex than in the absence of correlations.

Intuitively, this is due to the fact that for anticorrelated variables, an increase in one variable implies a decrease in the other one (different directional change): variables become more distant, thus more distinguishable in the Fisher-Rao information metric sense. Similarly, for positively correlated variables, an increase or decrease in one variable always predicts the same directional change for the second variable: variables do not become more distant, thus more distinguishable in the Fisher-Rao information metric sense. This may lead us to guess that in the presence of anticorrelations, motion on curved statistical manifolds via the Maximum Entropy updating methods becomes less complex.

3.3 Trivariate Gaussian statistical model

In this section we consider a Gaussian statistical model \mathcal{P} of the Eq. (10) when $n = 3$. In this case as well, in order to understand the asymptotic behaviour of the IGC in the presence of correlations between the micro-states, we make the minimal assumption that, given the random vector $X = (X_1, X_2, X_3)$ distributed according to a trivariate Gaussian, then $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \mathbb{E}(X_3) = \mu$ and $\mathbb{E}(X_1 - \mu)^2 = \mathbb{E}(X_2 - \mu)^2 = \mathbb{E}(X_3 - \mu)^2 = \sigma^2$. Therefore, the space of the parameters of \mathcal{P} is given by $\Theta = \{(\mu, \sigma) | \mu \in \mathbb{R}, \sigma > 0\}$.

The manifold $\mathcal{M} = (\Theta, g)$ changes its metric structure depending on the number of correlations between micro-variables, namely, one, two, or three. The covariance matrices corresponding to these cases read, modulo the congruence via a permutation matrix [17],

$$C_1 = \sigma^2 \begin{pmatrix} 1 & \rho & 0 \\ \rho & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad C_2 = \sigma^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & 0 \\ \rho & 0 & 1 \end{pmatrix}, \quad C_3 = \sigma^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}. \quad (27)$$

3.3.1 Case 1

First, we consider the trivariate Gaussian statistical model of Eq. (10) when $C \equiv C_1$. Then proceeding like in Section 3.2 we have $g = g_{11}d\mu^2 + g_{22}d\sigma^2$, where $g_{11} = \frac{3+\rho}{(1+\rho)\sigma^2}$

and $g_{22} = \frac{6}{\sigma^2}$. Also in this case we find that the sectional curvature of Eq. (5) is a negative function. Hence, as we state in Section 2, we may expect a decreasing (in time) behaviour of the information geometry complexity. Furthermore, we obtain the geodesics

$$\sigma(\tau) = \frac{2\sigma_0 \exp \left[\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right]}{1 + \exp \left[2\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right]}, \quad \mu(\tau) = -\frac{2\sigma_0 A_1}{\sqrt{\mathcal{A}(\rho)}} \frac{1}{1 + \exp \left[2\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right]}, \quad (28)$$

where $\mathcal{A}(\rho) = \frac{A_1^2(3+\rho)}{6(1+\rho)}$ and $A_1 \in \mathbb{R}$. We remark that $\mathcal{A}(\rho) > 0$ for all $\rho \in (-1, 1)$. Then, the volume (8) becomes

$$\text{vol} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau') \right] = \int \sqrt{\frac{6(3-4\rho)}{(1-2\rho^2)}} \frac{1}{\sigma^2} d\sigma d\mu = \frac{6A_1}{|A_1|} \exp \left[-\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right], \quad (29)$$

requiring $A_1 > 0$ for its positivity. Finally, using (29) in (7) we arrive at the asymptotic behaviour of the IGC

$$\widetilde{\text{vol}} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau) \right] \approx \left(\frac{6\sqrt{6}}{\sigma_0 A_1} \right) \sqrt{\frac{1+\rho}{3+\rho}} \frac{1}{\tau}. \quad (30)$$

Comparing (30) in the presence and in the absence of correlations yields

$$R_{\text{trivariate}}^{\text{weak}}(\rho) := \frac{\widetilde{\text{vol}} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau) \right]}{\widetilde{\text{vol}} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau) \right]_{\rho=0}} = \sqrt{3} \sqrt{\frac{1+\rho}{3+\rho}}, \quad (31)$$

where “weak” stands for low degree of connection in the lattice underlying the micro-variables

Notice that $R_{\text{trivariate}}^{\text{weak}}(\rho)$ is a monotonic increasing function of the argument $\rho \in (-1, 1)$.

3.3.2 Case 2

When the trivariate Gaussian statistical model of Eq.(10) has $C \equiv C_2$, the condition $C > 0$ constraints the correlation coefficient to be $\rho \in (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$. Proceeding again like in Section 3.2 we have $g = g_{11}d\mu^2 + g_{22}d\sigma^2$, where $g_{11} = \frac{3-4\rho}{(1-2\rho^2)\sigma^2}$ and $g_{22} = \frac{6}{\sigma^2}$. The sectional curvature of Eq. (5) is a negative function as well and so we may apply the arguments of Section 2 expecting a decreasing in time of the complexity. Furthermore, we obtain the geodesics

$$\sigma(\tau) = \frac{2\sigma_0 \exp \left[\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right]}{1 + \exp \left[2\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right]}, \quad \mu(\tau) = -\frac{2\sigma_0 A_1}{\sqrt{\mathcal{A}(\rho)}} \frac{1}{1 + \exp \left[2\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right]}, \quad (32)$$

where $\mathcal{A}(\rho) = \frac{A_1^2(3-4\rho)}{6(1-2\rho^2)}$ and $A_1 \in \mathbb{R}$. We remark that $\mathcal{A}(\rho) > 0$ for all $\rho \in (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$. Then, the volume (8) becomes

$$\text{vol} \left[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau') \right] = \int \sqrt{\frac{6(3-4\rho)}{(1-2\rho^2)}} \frac{1}{\sigma^2} d\sigma d\mu = \frac{6A_1}{|A_1|} \exp \left[-\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau \right]. \quad (33)$$

We have to set $A_1 > 0$ for the positivity of the volume (33), and using it in (7) we arrive at the asymptotic behaviour of the IGC

$$\widetilde{vol}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)] \approx \left(\frac{6\sqrt{6}}{\sigma_0 A_1}\right) \sqrt{\frac{1-2\rho^2}{3-4\rho}} \frac{1}{\tau}. \quad (34)$$

Then, comparing (34) in the presence and in the absence of correlations yields

$$R_{\text{trivariate}}^{\text{mildly weak}}(\rho) := \frac{\widetilde{vol}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)]}{\widetilde{vol}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)]_{\rho=0}} = \sqrt{3} \sqrt{\frac{1-2\rho^2}{3-4\rho}}, \quad (35)$$

where “mildly weak” stands for a lattice (underlying micro-variables) neither fully connected nor with minimal connection.

This is a function of the argument $\rho \in (-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ that attains the maximum $\sqrt{\frac{3}{2}}$ at $\rho = \frac{1}{2}$, while in the extrema of the interval $(-\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2})$ it tends to zero.

3.3.3 Case 3

Last, we consider the trivariate Gaussian statistical model of the Eq. (10) when $C \equiv C_3$. In this case, the condition $C > 0$ requires the correlation coefficient to be $\rho \in (-\frac{1}{2}, 1)$. Proceeding again like in Section 3.2 we have $g = g_{11}d\mu^2 + g_{22}d\sigma^2$, where $g_{11} = \frac{3}{(1+2\rho)\sigma^2}$ and $g_{22} = \frac{6}{\sigma^2}$. We find that the sectional curvature of Eq. (5) is a negative function; hence, we may expect a decreasing (in time) behaviour of the complexity. It follows the geodesics

$$\sigma(\tau) = \frac{2\sigma_0 \exp\left[\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau\right]}{1 + \exp\left[2\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau\right]}, \quad \mu(\tau) = -\frac{2\sigma_0 A_1}{\sqrt{\mathcal{A}(\rho)}} \frac{1}{1 + \exp\left[2\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau\right]}, \quad (36)$$

where $\mathcal{A}(\rho) = \frac{A_1^2}{2(1+2\rho)}$ and $A_1 \in \mathbb{R}$. We note that $\mathcal{A}(\rho) > 0$ for all $\rho \in (-\frac{1}{2}, 1)$. Using (36), we compute

$$vol[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau')] = \int \frac{3\sqrt{2}}{\sqrt{(1+2\rho)}} \frac{1}{\sigma^2} d\sigma d\mu = \frac{6\sqrt{2}A_1}{|A_1|} \exp\left[-\sigma_0 \sqrt{\mathcal{A}(\rho)} \tau\right]. \quad (37)$$

Also in this case we need to assume $A_1 > 0$ to have positive volume. Finally, substituting Eq. (37) into Eq. (7), the asymptotic behaviour of the IGC results

$$\widetilde{vol}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)] \approx \left(\frac{12}{\sigma_0 A_1}\right) \sqrt{1+2\rho} \frac{1}{\tau}. \quad (38)$$

The comparison of (38) in the presence and in the absence of correlations yields

$$R_{\text{trivariate}}^{\text{strong}}(\rho) := \frac{\widetilde{vol}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)]}{\widetilde{vol}[\mathcal{D}_{\Theta}^{(\text{geodesic})}(\tau)]_{\rho=0}} = \sqrt{1+2\rho}, \quad (39)$$

where “strong” stands for a fully connected lattice underlying the (three) micro-variables. We remark the latter ratio is a monotonically increasing function of the argument $\rho \in (-\frac{1}{2}, 1)$.

The behaviors of $R(\rho)$ of Eqs.(26), (31), (35), (39) are reported in Fig.1.

The *non-monotonic* behavior of the ratio $R_{\text{trivariate}}^{\text{mildly weak}}(\rho)$ in Eq. (35) corresponds to the information geometric complexities for the mildly weak connected three-dimensional lattice. Interestingly, the growth stops at a critical value $\rho_{\text{peak}} = \frac{1}{2}$ at which $R_{\text{trivariate}}^{\text{mildly weak}}(\rho_{\text{peak}}) = R_{\text{bivariate}}^{\text{strong}}(\rho_{\text{peak}})$. From Eq. (26), we conclude that entropic inferences on three Gaussian distributed micro-variables on a fully connected lattice is carried out in a more efficient manner when the two micro-variables are negatively correlated. Instead, when such micro-variables are positively correlated, macroscopic predictions become more complex than in the absence of correlations. Furthermore, the ratio $R_{\text{trivariate}}^{\text{strong}}(\rho)$ of the information geometric complexities for this fully connected three-dimensional lattice increases in a *monotonic* fashion. These conclusions are similar to those presented for the bivariate case. However, there is a key-feature of the IGC to emphasize when passing from the two-dimensional to the three-dimensional manifolds associated with fully connected lattices: the effects of negative-correlations and positive-correlations are *amplified* with respect to the respective absence of correlations scenarios,

$$\frac{R_{\text{trivariate}}^{\text{strong}}(\rho)}{R_{\text{bivariate}}^{\text{strong}}(\rho)} = \sqrt{\frac{1+2\rho}{1+\rho}}, \quad (40)$$

where $\rho \in (-\frac{1}{2}, 1)$.

Specifically, carrying out entropic inferences on the higher-dimensional manifold in the presence of anti-correlations, that is for $\rho \in (-\frac{1}{2}, 0)$, is less complex than on the lower-dimensional manifold as evident from Eq. (40). The vice-versa is true in the presence of positive-correlations, that is for $\rho \in (0, 1)$.

4 Concluding remarks

In summary, we considered low dimensional Gaussian statistical models (up to a trivariate model) and have investigated their dynamical (temporal) complexity. This has been quantified by the volume of geodesics for parameters characterizing the probability distribution functions.

We uncover that in order to have a reduction in time of the complexity one has to consider both mean and variance as macro-variables. This leads to different topological structures of the parameter space in (11); in particular, we have to consider at least a 2-dimensional manifold in order to have effects such as a power law decay of the complexity. Hence, the minimal hypothesis in a multivariate Gaussian model consists in considering all mean values equal and all covariances equal. In such a case, however, the complexity shows interesting features depending on the correlation among micro-variables (as summarized in Fig.1). For a trivariate model with only two correlations the information geometric complexity ratio exhibits a non monotonic behaviour in ρ (correlation parameter) taking zero value at the extrema

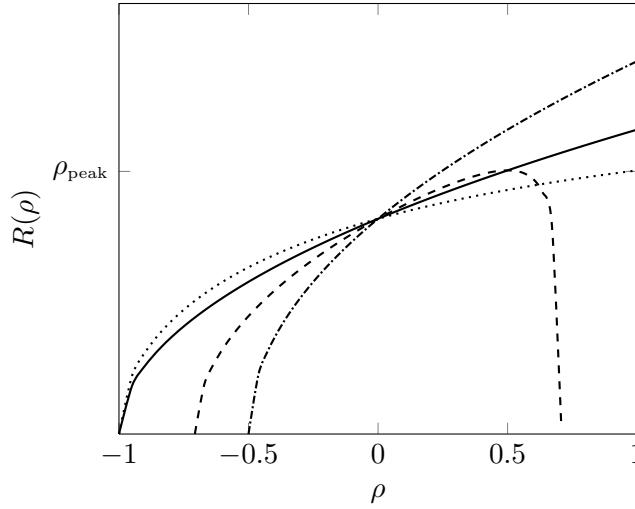


Figure 1: Ratio $R(\rho)$ of volumes vs. degree of correlations ρ . Solid line refers to $R_{\text{bivariate}}^{\text{strong}}(\rho)$; Dotted line refers to $R_{\text{trivariate}}^{\text{weak}}(\rho)$; Dashed line refers to $R_{\text{trivariate}}^{\text{mildly weak}}(\rho)$; Dash-dotted refers to $R_{\text{trivariate}}^{\text{strong}}(\rho)$.

of the range of ρ . In contrast to closed configurations (bivariate and trivariate models with all micro-variables correlated each other) the complexity ratio exhibits a monotonic behaviour in terms of the correlation parameter. The fact that in such a case this ratio cannot be zero at the extrema of the range of ρ is reminiscent of the geometric frustration phenomena that occurs in the presence of loops [11].

Specifically, recall that a geometrically frustrated system cannot simultaneously minimize all interactions because of geometric constraints [11, 18]. For example, geometric frustration can occur in an Ising model which is an array of spins (for instance, atoms that can take states ± 1) that are magnetically coupled to each other. If one spin is, say, in the $+1$ state then it is energetically favorable for its immediate neighbours to be in the same state in the case of a ferromagnetic model. On the contrary, in antiferromagnetic systems, nearest neighbor spins want to align in opposite directions. This rule can be easily satisfied on a square. However, due to geometrical frustration, it is not possible to satisfy it on a triangle: for an antiferromagnetic triangular Ising model, any three neighboring spins are frustrated. Geometric frustration in triangular Ising models can be observed by considering spin configurations with total spin $J = \pm 1$ and analyzing the fluctuations in energy of the spin system as a function of temperature. There is no peak at all in the standard deviation of the energy in the case $J = -1$, and a monotonic behaviour is recorded. This indicates that the antiferromagnetic system does not have a phase transition to a state with long-range order. Instead, in the case $J = +1$, a peak in the energy fluctuations emerges. This significant change in the behaviour of energy fluctuations as a function of temperature in triangular configurations of spin systems is a signature of the presence of frustrated interactions in the system [19].

In this article, we observe a significant change in the behaviour of the information geometric complexity ratios as a function of the correlation coefficient in the trivariate Gaussian statistical models. Specifically, in the fully connected trivariate

case, no peak arises and a monotonic behaviour in ρ of the information geometric complexity ratio is observed. In the mildly weak connected trivariate case, instead, a peak in the information geometric complexity ratio is recorded at $\rho_{\text{peak}} \geq 0$. This dramatic disparity of behaviour can be ascribed to the fact that when carrying out statistical inferences with positively correlated Gaussian random variables, the maximum entropy favorable scenario is incompatible with these working hypothesis. Thus, the system appears frustrated.

These considerations lead us to conclude that we have uncovered a very interesting information geometric resemblance of the more standard geometric frustration effect in Ising spin models. However, for a conclusive claim of the existence of an information geometric analog of the frustration effect, we feel we have to further deepen our understanding. A forthcoming research project along these lines will be a detailed investigation of both arbitrary triangular and square configurations of correlated Gaussian random variables where we take into consideration both the presence of different intensities and signs of pairwise interactions ($\rho_{ij} \neq \rho_{ik}$ if $j \neq k$, $\forall i$).

Acknowledgments

D.F. and S.M. acknowledge the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under the FET-Open grant agreement TOPDRIM, number FP7-ICT-318121.

References

- [1] D. F. Feldman, J. P. Crutchfield, *Measures of Statistical Complexity: Why?*, Physics Letters A, Vol. **238** (1998).
- [2] A. N. Kolmogorov, *A new metric invariant of transitive dynamical systems and of automorphism of Lebesgue spaces*, Doklady Akademii Nauk SSSR **119** (1958).
- [3] A. Caticha, Entropic dynamics, in: R.L. Fry(Ed.), Bayesian Inference and Maximum Entropy Methods in Science and Engineering, in AIP Conf. Proc., **617**, 302 (2002).
- [4] A. Caticha and R. Preuss, *Maximum entropy and data analysis: Entropic prior distributions*, Physical Reviews E **70** (2004).
- [5] S. Amari, and H. Nagaoka, *Methods of Information Geometry*, Oxford University Press (2000).
- [6] C. Cafaro, *Works on an information geometrodynamical approach to chaos*, Chaos Solitons and Fractals **41**, 886 (2009).
- [7] M. Pettini, *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer-Verlag (2007).

- [8] J. L. Lebowitz, *Microscopic Dynamics and Macroscopic Laws*, Annals of the New York Academy of Sciences **373**, 220 (1981).
- [9] T. Shibata et al., *Noiseless Collective Motion out of Noisy Chaos*, Phys. Rev. Lett. **82**, 4424 (1999).
- [10] S.A. Ali, C. Cafaro, D.-H. Kim, and S. Mancini, *The effect of the microscopic correlations on the information geometric complexity of Gaussian statistical models*, Physica A **389**, 3117 (2010).
- [11] J. F. Sadoc and R. Mosseri, *Geometrical Frustration*, Cambridge University Press (2006).
- [12] J. M. Lee *Riemannian Manifolds: an Introduction to Curvature*, Springer (1997).
- [13] M. P. do Carmo, *Riemannian Geometry*, Birkäuser (1992).
- [14] C. Cafaro and S. A. Ali, *Jacobi fields on statistical manifolds of negative curvature*, Physica D: Nonlinear Phenomena **234**, 70 (2007).
- [15] C. Cafaro, A. Giffin, S. A. Ali, and D.-H. Kim, *Reexamination of an information geometric construction of entropic indicators of complexity*, Appl. Math. Comput. **217**, 2944 (2010).
- [16] C. Cafaro and S. Mancini, *Quantifying the complexity of geodesic paths on curved statistical manifolds through information geometric entropies and Jacobi fields*, Physica D: Nonlinear Phenomena **240**, 607 (2011).
- [17] D. Felice, S. Mancini, M. Pettini, *Quantifying Networks Complexity from Information Geometry Viewpoint*, arXiv:1310.7825 math-physics.
- [18] R. Moessner and A. P. Ramirez, *Geometrical Frustration*, Physics Today **59**, 24 (2006).
- [19] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press (2003)